

# Machine Learning & Data Mining

## Lecture 4

Corina Besliu

Technical University of Moldova

November 12, 2021



# Outline

- Decision Tree Regression Model
- Random Forest Regression Model

# Outline

- Decision Tree Regression Model
- Random Forest Regression Model

# Decision Tree Regressions

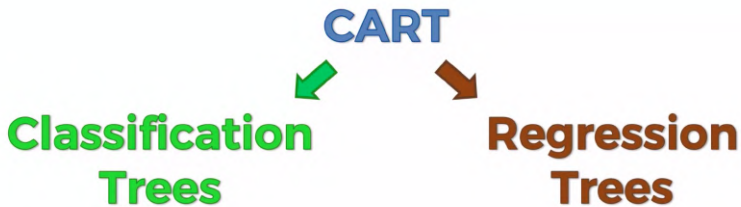
# Decision Tree Intuition

---

**CART**

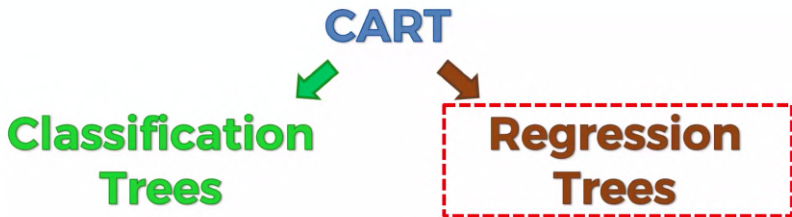
# Decision Tree Intuition

---

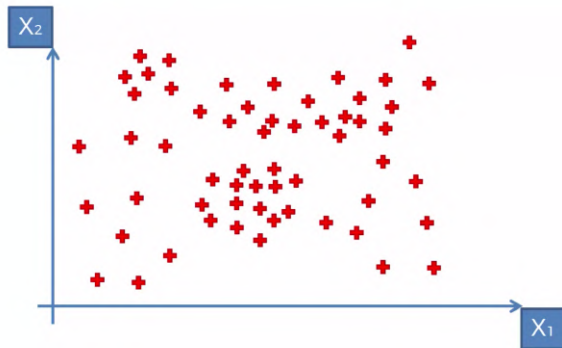


# Decision Tree Intuition

---

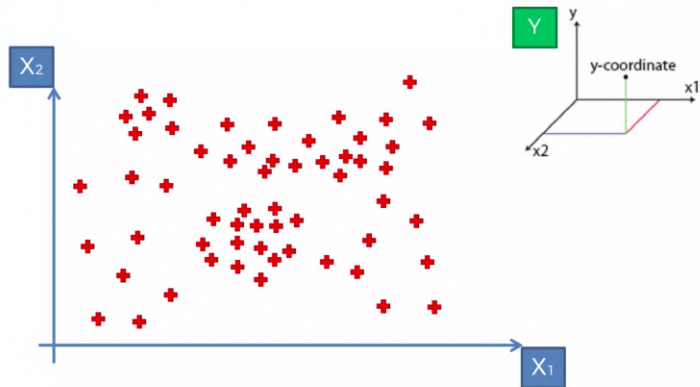


# Decision Tree Intuition

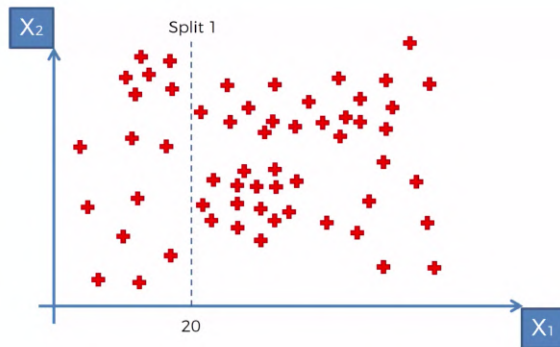




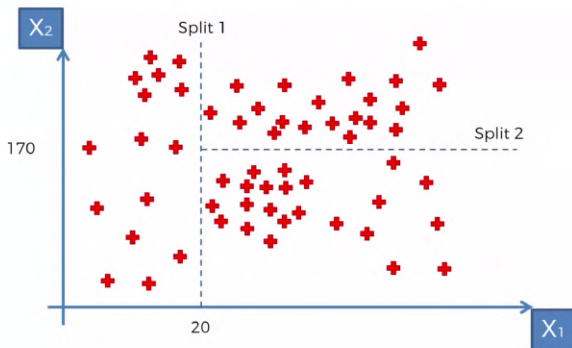
# Decision Tree Intuition



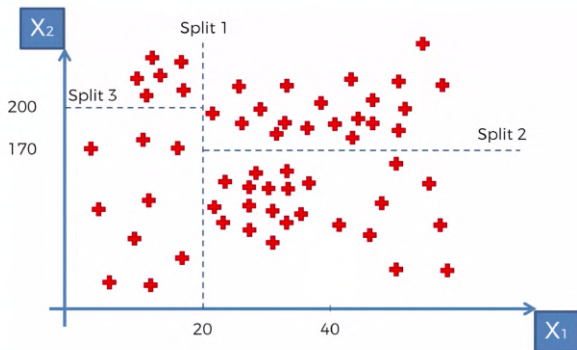
# Decision Tree Intuition



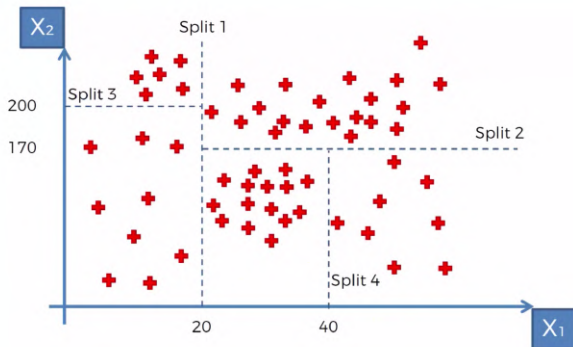
# Decision Tree Intuition



# Decision Tree Intuition



# Decision Tree Intuition

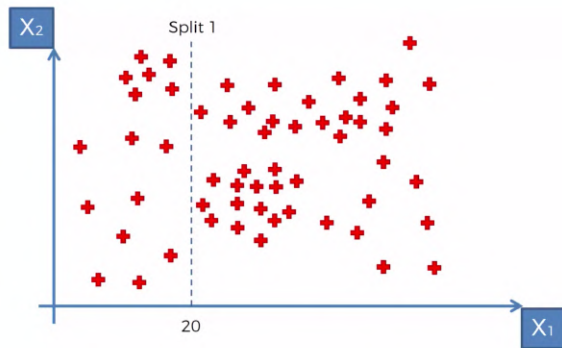


# Decision Tree Intuition

---

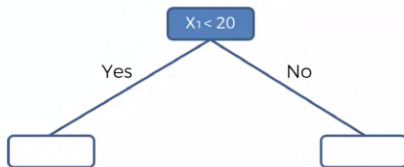
**Rewind...**

# Decision Tree Intuition



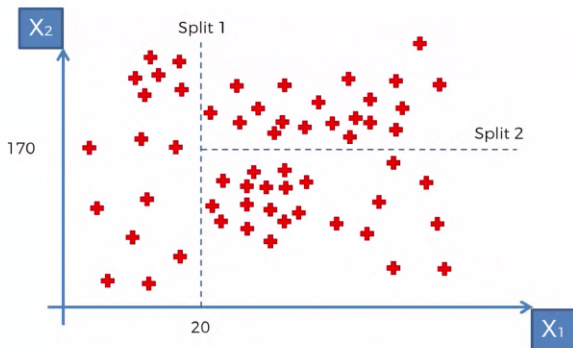
# Decision Tree Intuition

---

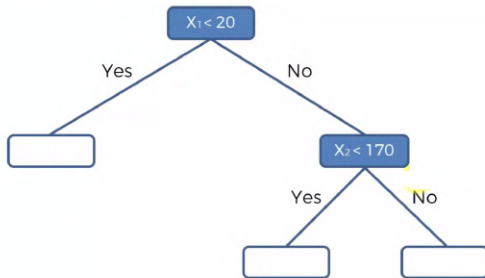




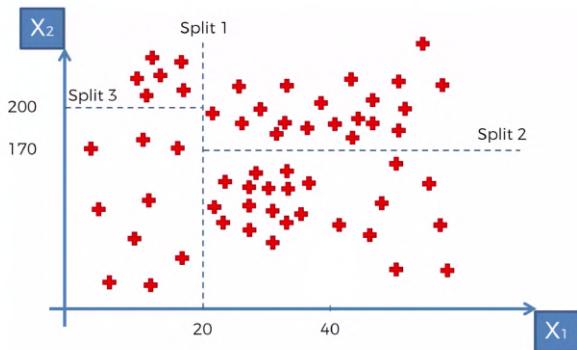
# Decision Tree Intuition



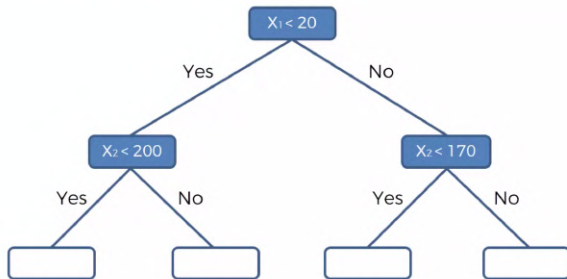
# Decision Tree Intuition



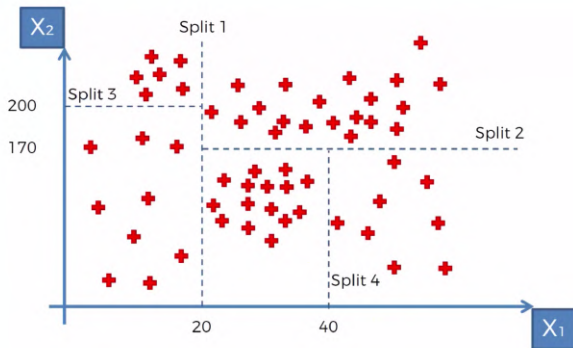
# Decision Tree Intuition



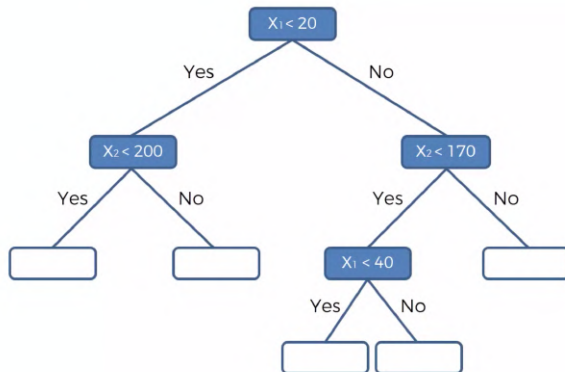
# Decision Tree Intuition



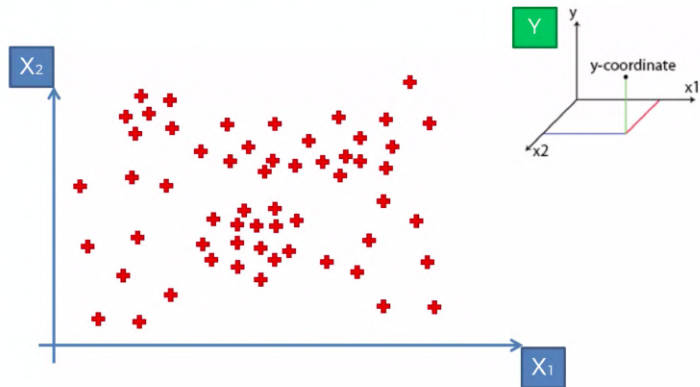
# Decision Tree Intuition



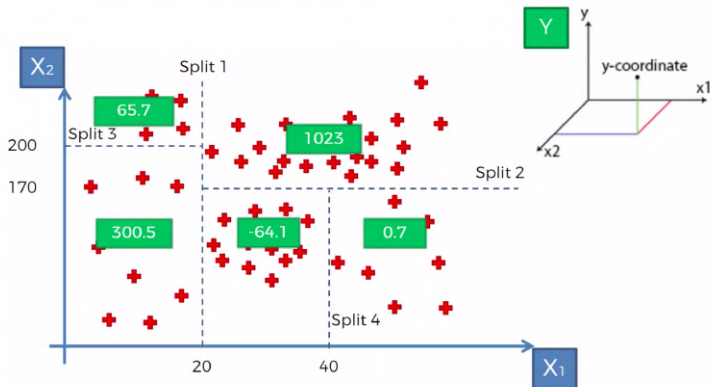
# Decision Tree Intuition



# Decision Tree Intuition

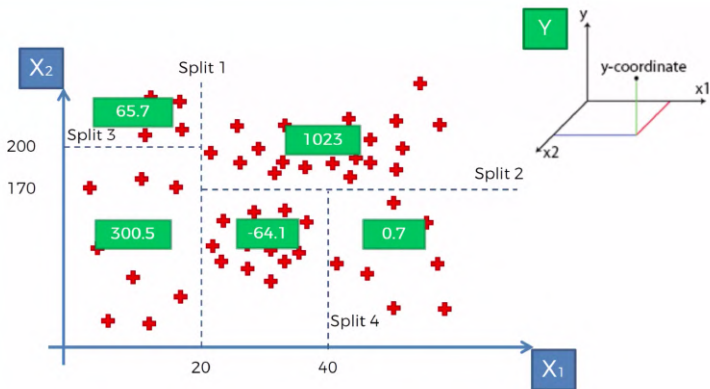


# Decision Tree Intuition



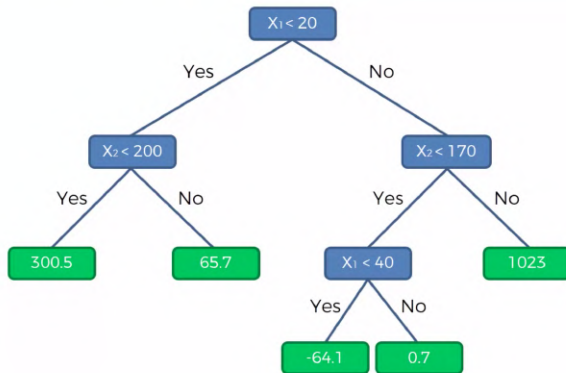


# Decision Tree Intuition



Example :  $x_1 = 50$   $x_2 = 10$

# Decision Tree Intuition



# Making the Splits

**How are the splits done?**

<https://www.youtube.com/watch?v=g9c66TUylZ4>

# Random Forest Regression

# Random Forest Intuition

---

## Ensemble Learning

# Random Forest Intuition

---

STEP 1: Pick at random  $K$  data points from the Training set.

# Random Forest Intuition

---

STEP 1: Pick at random  $K$  data points from the Training set.



STEP 2: Build the Decision Tree associated to these  $K$  data points.

# Random Forest Intuition

---

STEP 1: Pick at random  $K$  data points from the Training set.



STEP 2: Build the Decision Tree associated to these  $K$  data points.



STEP 3: Choose the number  $N_{tree}$  of trees you want to build and repeat STEPS 1 & 2.



# Random Forest Intuition

---

STEP 1: Pick at random  $K$  data points from the Training set.



STEP 2: Build the Decision Tree associated to these  $K$  data points.



STEP 3: Choose the number  $N_{tree}$  of trees you want to build and repeat STEPS 1 & 2



STEP 4: For a new data point, make each one of your  $N_{tree}$  trees predict the value of  $Y$  to for the data point in question, and assign the new data point the average across all of the predicted  $Y$  values.

# Standardization & Normalization

## Feature Scaling

Standardisation	Normalisation
$x_{\text{stand}} = \frac{x - \text{mean}(x)}{\text{standard deviation}(x)}$	$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$

# Standardization & Normalization

$$\mu = 0 \quad \sigma = 1$$

When to use one versus the other?

## Standardization

- When the distribution of data is unknown
- When model is sensitive to different scales (SVR)

## Normalization

- When the distribution of features is close to normal
- When model assumes normal distribution (e.g. Linear Regression, Logistic Regression, etc.)

# Standardization & Normalization

$$\mu = 0 \quad \sigma = 1$$

**When to use one versus the other?**

## Standardization

- When the distribution of data is unknown
- When model is sensitive to different scales (SVR)

## Normalization

- When the distribution of features is close to normal
- When model assumes normal distribution (e.g. Linear Regression, Logistic Regression, etc.)

# Standardization & Normalization

$$\mu = 0 \quad \sigma = 1$$

**When to use one versus the other?**

## Standardization

- When the distribution of data is unknown
- When model is sensitive to different scales (SVR)

## Normalization

- When the distribution of features is close to normal
- When model assumes normal distribution (e.g. Linear Regression, Logistic Regression, etc.)

# Standardization & Normalization

$$\mu = 0 \quad \sigma = 1$$

**When to use one versus the other?**

## **Standardization**

- When the distribution of data is unknown
- When model is sensitive to different scales (SVR)

## **Normalization**

- When the distribution of features is close to normal
- When model assumes normal distribution (e.g. Linear Regression, Logistic Regression, etc.)

# Let's get Started!

**Access Google Colaboratory through your Gmail account**